

The Least Degraded and the Least Upgraded Channel with respect to a Channel Family

Wei Liu, S. Hamed Hassani, and Rüdiger Urbanke
School of Computer and Communication Sciences
EPFL, Switzerland

Email: {wei.liu, seyedhamed.hassani, rudiger.urbanke}@epfl.ch

Abstract—Given a family of binary-input memoryless output-symmetric (BMS) channels having a fixed capacity, we derive the BMS channel having the highest (resp. lowest) capacity among all channels that are degraded (resp. upgraded) with respect to the whole family. We give an explicit characterization of this channel as well as an explicit formula for the capacity of this channel.

I. INTRODUCTION

Channel ordering plays an important role in analyzing the asymptotic behavior of iterative coding systems over binary-input memoryless output-symmetric (BMS) channels. We say that two BMS channels are *ordered* by a real-valued parameter, if the “worse” channel has a larger parameter than the “better” channel. E.g., we might consider as a real-valued parameter the bit-error probability, the capacity, or perhaps the Battacharyya parameter.

One particularly useful way of imposing an order is to consider the notion of *degradation* [1]. The main reason that this order is so useful is that many important functionals and kernels (defined later) associated with BMS channels either preserve or reserve the ordering induced by degradation. In addition, the order of degradation is preserved under many natural operations one might perform on channels.

Instead of considering a single BMS channel, we are interested in a family of BMS channels sharing certain properties. The cardinality of the family can be either finite or infinite. A particularly important example for our purpose is the family of BMS channels having fixed capacity c . This family is denoted by $\{\text{BMS}(c)\}$. The question we address is the following. Given the family of all BMS channels of capacity c , can we find the “best” degraded channel and the “worst” upgraded channel with respect to this family. I.e., can we find the channel that is degraded with respect to all members in the family and has the highest capacity, as well as the channel that is upgraded with respect to all members of the family and has the lowest capacity.

Determining these channels is both natural and fundamental. To mention just one possible application, assume that we want to construct a polar code [2] which works for all channels of a particular capacity. If we can find a channel which is degraded with respect to all elements in the family then a polar code designed for this channel will work a fortiori for all elements of the original family. We then say that this polar code is *universal*. This simple bound is somewhat too crude to construct a good universal polar code. But we can apply the same argument not to the original family but after a few steps of the polarization process. The more steps we perform

the better the overall code will be (i.e., it will have higher capacity) and the closer we will get to an optimal construction. In what follows we will not be concerned with applications but we will only be interested in describing these two extreme channels given a channel family. As we will see, we will be able to give an explicit answer to the question.

A. Channel Model

A binary-input memoryless output-symmetric (BMS) channel has an input X taking values over an input alphabet \mathcal{X} , an output Y taking values over an output alphabet \mathcal{Y} , and a transition probability $p_{Y|X}(y|x)$ for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Throughout the paper we take the input alphabet \mathcal{X} to be $\{\pm 1\}$ and the output alphabet \mathcal{Y} to be a subset of $\bar{R} \triangleq [-\infty, +\infty]$.

Given a BMS channel a , an alternative description of the channel is the L -distribution of the random variable

$$L = l(Y) \triangleq \ln \frac{p_{Y|X}(Y|1)}{p_{Y|X}(Y|-1)}.$$

A closely related random variable is

$$D = d(Y) \triangleq \tanh\left(\frac{l(Y)}{2}\right) = \frac{1 - e^{-l(Y)}}{1 + e^{-l(Y)}},$$

whose distribution is called a D -distribution, conditioned on $X = 1$. Moreover, denote by $|D|$ the absolute value of the random variable D . Then, the distribution of $|D|$ is termed a $|D|$ -distribution and the associated density is called a $|D|$ -density. Given a BMS channel a , we denote by $|\mathfrak{A}|$ and $|\mathfrak{a}|$ the associated $|D|$ -distribution and $|D|$ -density, respectively. The $|D|$ -density of a *discrete* BMS channels is of the form

$$|\mathfrak{a}|(x) = \sum_{i=1}^n \alpha_i \delta(x - x_i), \quad (1)$$

where $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^n \alpha_i = 1$, and $0 \leq x_1 < x_2 < \dots < x_n \leq 1$; while $\delta(\cdot)$ is the (Dirac) delta function and $n \in \mathbb{N}_+$ is finite or countably infinite.

B. Functionals of Densities

The *capacity* of a BMS channel can be defined as a linear functional of its $|D|$ -density [1]. Formally, the capacity $C(|\mathfrak{a}|)$ in bits per channel use of a BMS channel with D -density $|\mathfrak{a}|$ is defined as

$$C(|\mathfrak{a}|) \triangleq \int_0^1 |\mathfrak{a}|(x) (1 - h(x)) dx,$$

where $h(x) \triangleq h_2((1-x)/2)$, for $x \in [0, 1]$, and $h_2(x) \triangleq -x \log_2(x) - (1-x) \log_2(1-x)$ is the *binary entropy function*. The *entropy functional* is defined as $H(|a|) \triangleq 1 - C(|a|)$, and the entropy of a discrete BMS channel is $\sum_{i=1}^n \alpha_i h(x_i)$.

Another important functional is the *Battacharyya parameter* associated with the $|D|$ -density $|a|$. It is defined as

$$\mathfrak{B}(|a|) \triangleq \int_0^1 |a|(x) \sqrt{1-x^2} dx.$$

If a BMS channel has a $|D|$ -density of the form in (1), the Battacharyya parameter is given by $\sum_{i=1}^n \alpha_i \sqrt{1-x_i^2}$.

C. Degradedness

Denote by $p_{Y|X}(y|x)$ and $p_{Z|X}(z|x)$ the transition probabilities of two BMS channels a and a' , respectively. Let \mathcal{X} be the common input alphabet, and denote by \mathcal{Y} and \mathcal{Z} the output alphabets, respectively. We say that a' is (*stochastically*) *degraded* with respect to a if there exists a memoryless channel with transition probability $p_{Z|Y}(z|y)$ such that

$$p_{Z|X}(z|x) = \sum_{y \in \mathcal{Y}} p_{Z|Y}(z|y) p_{Y|X}(y|x),$$

for all $x \in \mathcal{X}$ and all $z \in \mathcal{Z}$. Conversely, the channel a is said to be (*stochastically*) *upgraded* with respect to the channel a' . The following lemma gives an equivalent characterization of degradedness and upgradedness [1].

Lemma 1. Consider two BMS channels a and a' with the corresponding $|D|$ -distributions $|a|$ and $|a'|$, respectively. The following statements are equivalent:

- (i) a' is degraded with respect to a ;
- (ii) $\int_0^1 f(x) d|a|(x) \leq \int_0^1 f(x) d|a'|(x)$, for all f that are non-increasing and convex- \cap on $[0, 1]$;
- (iii) $\int_z^1 |a|(x) dx \leq \int_z^1 |a'|(x) dx$, for all $z \in [0, 1]$.

II. LEAST DEGRADED CHANNEL

Now recall that the entropy functional $H(|a|)$ associated with the $|D|$ -density $|a|$ has $h_2((1-x)/2)$ as its kernel in the $|D|$ -domain. This kernel is non-increasing and convex- \cap on $[0, 1]$, and therefore condition (ii) in Lemma 1 shows that a degraded channel has a larger entropy, and thus lower capacity, than the original channel. Similarly, a degraded channel has a larger probability of error and larger Battacharyya parameter than the original channel.

Property 2. Note that there are many equivalent definitions for the *least degraded* channel. One particularly insightful characterization is that the least degraded channel is the unique channel which is degraded with respect to all elements of the family $\{\text{BMS}(c)\}$ and is upgraded with respect to any other such channel. Therefore it has the highest capacity, lowest error probability and lowest Battacharyya parameter.

The integral of $|a|(x)$ from z to 1, as suggested by Lemma 1, is important for characterizing degradedness among BMS channels. Therefore, for a BMS channel a , define

$$\Lambda_a(z) \triangleq \int_z^1 |a|(x) dx. \quad (2)$$

Note that $\Lambda_a(z)$ is decreasing on $[0, 1]$, and since $|a|(x)$ is increasing in x , it follows that $\Lambda_a(z)$ is a convex- \cap function of z . Moreover, if the BMS channel a is discrete, the function $\Lambda_a(z)$ has a simpler form given by

$$\Lambda_a(z) = \sum_{i=1}^n \alpha_i (1 - \max\{z, x_i\}). \quad (3)$$

Condition (iii) of Lemma 1 shows that, in order to find the least degraded channel with respect to $\{\text{BMS}(c)\}$, we take the maximum value of $\Lambda_a(z)$ among all BMS channels $a \in \{\text{BMS}(c)\}$, for each fixed $z \in [0, 1]$. E.g., define

$$\bar{\Lambda}(z) \triangleq \max\{\Lambda_a(z) : a \in \{\text{BMS}(c)\}\}, \quad (4)$$

for every $z \in [0, 1]$. Taking the convex- \cap envelope of $\bar{\Lambda}(z)$ then characterizes the desired channel. The following theorem characterizes $\bar{\Lambda}(z)$ exactly.

Theorem 3. Consider the family of BMS channels $\{\text{BMS}(c)\}$ of capacity c , $0 < c < 1$. Then,

$$\bar{\Lambda}(z) = \begin{cases} \frac{(1-c)(1-z)}{h(z)} & \text{if } z \in [0, 1-2\epsilon_{\text{bsc}}], \\ 1-z & \text{if } z \in [1-2\epsilon_{\text{bsc}}, 1], \end{cases}$$

where $\epsilon_{\text{bsc}} \in (0, \frac{1}{2})$ is the solution of $1 - h_2(\epsilon_{\text{bsc}}) = c$.

Proof: First, it is clear that $\sum_{i=1}^n \alpha_i (1 - \max\{z, x_i\}) \leq \sum_{i=1}^n \alpha_i (1 - z) = 1 - z$, and for $z \in [1 - 2\epsilon_{\text{bsc}}, 1]$ this value, $1 - z$, is achieved if the underlying BMS channel is the BSC with crossover probability equal to ϵ_{bsc} .

Now for any fixed $z \in [0, 1 - 2\epsilon_{\text{bsc}}]$, assume that $\bar{\Lambda}(z)$ is achieved by the BMS channel d , e.g., $\bar{\Lambda}(z) = \Lambda_d(z)$. We claim that d does not have any probability mass in the interval $(z, 1)$. We show this by contradiction. Suppose that there exists a probability mass α_0 at $x_0 \in (z, 1)$. Define $\rho \triangleq h(x_0)/h(z)$. It is clear that $\rho \in (0, 1)$, since the function $h(\cdot)$ is decreasing and convex- \cap on $[0, 1]$. The definition of ρ gives

$$\alpha_0 h(x_0) = \alpha_0 \rho h(z) + \alpha_0 (1 - \rho) h(1), \quad (5)$$

which means that, instead of putting the single probability mass α_0 at x_0 , we can split α_0 into two masses, $\alpha_0 \rho$ and $\alpha_0 (1 - \rho)$, and put these two masses at z and 1, respectively, without changing the entropy and thus the capacity. Canceling α_0 on both sides of (5) and using the fact that $h(\cdot)$ is decreasing and convex- \cap on $[0, 1]$, we have

$$h(x_0) < h(\rho z + 1 - \rho),$$

which is equivalent to $1 - x_0 < \rho(1 - z)$. Now notice that, since $z < x_0 < 1$, the term corresponding to x_0 that contributes to $\Lambda_d(z)$ is equal to $\alpha_0(1 - x_0)$. But

$$\alpha_0(1 - x_0) < \alpha_0 \rho(1 - z) = \alpha_0 \rho(1 - z) + \alpha_0(1 - \rho)(1 - 1),$$

which means that the value of $\Lambda_d(z)$ can be increased by the splitting operation mentioned above. This, however, contradicts the assumption that $\Lambda_d(z)$ is the maximum value at z . If there exists some probability masses on the interval $[0, z]$, we can add these masses to the probability mass at z and delete the original masses, without changing $\Lambda_d(z)$ and without violating the entropy (or capacity) constraint. Thus, the channel d has

probability masses at points z and 1 only, and the probability mass at z is equal to $(1 - c)/h(z)$, completing the proof. ■

Recall that the function Λ associated to a BMS channel is convex- \cap on $[0, 1]$. Thus, taking the convex- \cap envelope of $\bar{\Lambda}(z)$ gives the Λ function, call it $\Lambda^*(z)$, of the least degraded channel with respect to the whole channel family $\{\text{BMS}(c)\}$.

Corollary 4. The least degraded channel is characterized by

$$\Lambda^*(z) = \begin{cases} 1 - c - \frac{1 - c - 2\epsilon_{\text{bsc}}}{1 - 2\epsilon_{\text{bsc}}}z & \text{if } z \in [0, 1 - 2\epsilon_{\text{bsc}}), \\ 1 - z & \text{if } z \in [1 - 2\epsilon_{\text{bsc}}, 1]. \end{cases}$$

Once having this characterization, we can derive the exact formula of the capacity of the least degraded channel.

Theorem 5. Given a family of BMS channels of capacity c , the capacity in bits per channel use of the least degraded channel is given by

$$C^* = \frac{c^2}{1 - 2\epsilon_{\text{bsc}}}. \quad (6)$$

Proof: Recall that the entropy can be expressed in the following alternative form,

$$H(|a|) = \int_0^1 \frac{1}{\ln 2(1 - z^2)} \left(\int_z^1 |\mathfrak{A}|(x) \right) dz. \quad (7)$$

Inserting the formula for $\Lambda^*(z)$ into (7) and integrating over z gives the desired result. ■

III. LEAST UPGRADED CHANNEL

Property 6. The *least upgraded* channel is the unique channel which is upgraded with respect to all elements of the family $\{\text{BMS}(c)\}$ and is degraded with respect to any other such channel. Therefore this channel has the lowest capacity, highest error probability and highest Battacharyya parameter.

In order to specify the least upgraded channel, we first notice the following lemma.

Lemma 7. For any BMS channel associated with the $|D|$ -density of the form in (1) and having capacity c , $0 < c < 1$, we have $1 - 2\epsilon_{\text{bsc}} \leq x_n \leq 1$.

Proof: Assume on the contrary that $x_n < 1 - 2\epsilon_{\text{bsc}}$. Then, the monotonicity property of $h(\cdot)$ gives

$$h(x_n) > h(1 - 2\epsilon_{\text{bsc}}) = 1 - c,$$

which is equivalent to $\sum_{i=1}^n \alpha_i h(x_i) > 1 - c$, contradicting the assumption that the entropy is $1 - c$. ■

Now, for each fixed $z \in [0, 1]$, we take the minimum value of $\Lambda_a(z)$ among all BMS channels $a \in \{\text{BMS}(c)\}$, e.g., define

$$\underline{\Lambda}(z) \triangleq \min \{\Lambda_a(z) : a \in \{\text{BMS}(c)\}\}. \quad (8)$$

Suppose that, for any fixed $z \in [0, 1]$, the minimum value is achieved by the channel u , i.e., $\underline{\Lambda}(z) = \Lambda_u(z)$. For channel u , the number of probability masses on the interval $[z, 1]$ is characterized by the following lemma.

Lemma 8. The channel u , which achieves the minimum value $\underline{\Lambda}(z)$ at z , has at most one probability mass on $[z, 1]$.

Proof: Assume there are two probability masses α_- and α_+ at $x_- \in [z, 1]$ and $x_+ \in [z, 1]$, respectively. Then, there exists a point \tilde{x} on $[x_-, x_+]$ such that

$$\alpha_- h(x_-) + \alpha_+ h(x_+) = (\alpha_- + \alpha_+) h(\tilde{x}). \quad (9)$$

Dividing both sides of (9) by $(\alpha_- + \alpha_+)$ and using the convex- \cap property of $h(\cdot)$ gives

$$h(\tilde{x}) < h\left(\frac{\alpha_-}{\alpha_- + \alpha_+}x_- + \frac{\alpha_+}{\alpha_- + \alpha_+}x_+\right),$$

which means that $\alpha_-x_- + \alpha_+x_+ < (\alpha_- + \alpha_+)\tilde{x}$. Since the terms corresponding to x_- and x_+ that contribute to $\Lambda_u(z)$ is equal to $\alpha_-(1 - x_-) + \alpha_+(1 - x_+)$, we have

$$\begin{aligned} & \alpha_-(1 - x_-) + \alpha_+(1 - x_+) \\ &= (\alpha_- + \alpha_+) \left(1 - \frac{\alpha_-}{\alpha_- + \alpha_+}x_- - \frac{\alpha_+}{\alpha_- + \alpha_+}x_+ \right) \\ &> (\alpha_- + \alpha_+)(1 - \tilde{x}), \end{aligned}$$

which shows that by combining two probability masses α_- and α_+ into a single probability mass $(\alpha_- + \alpha_+)$ at position \tilde{x} , the value of $\Lambda_u(z)$ is decreased, contradicting the assumption that $\Lambda_u(z)$ is minimal at point z . ■

Lemma 7 and Lemma 8 show that the channel u has at most two probability masses on the interval $[0, 1]$. If it has only one probability mass, it is in fact a BSC. Otherwise, in general, there are two probability masses, call them γ and $1 - \gamma$, on the intervals $[0, z]$ and $[1 - 2\epsilon_{\text{bsc}}, 1]$, respectively. Denote by \tilde{x} and \hat{x} the positions of these two masses, respectively.

Lemma 9. Either $\tilde{x} = 0$ holds or $\hat{x} = 1$ holds, or $\tilde{x} = 0$ and $\hat{x} = 1$ hold simultaneously.

Proof: Suppose on the contrary that $\tilde{x} \neq 0$ and $\hat{x} \neq 1$. Then, consider decreasing \tilde{x} by ζ , where $\zeta \in \mathbb{R}_+$ is sufficiently small. Assume that \hat{x} is increased by δ correspondingly, where $\delta \in \mathbb{R}_+$ is sufficiently small. First-order Taylor expansion gives

$$\begin{aligned} h(\tilde{x} - \zeta) &= h(\tilde{x}) - \zeta h'(\tilde{x}) + \mathcal{O}(\zeta^2), \\ h(\hat{x} + \delta) &= h(\hat{x}) + \delta h'(\hat{x}) + \mathcal{O}(\delta^2). \end{aligned}$$

Eliminating second and higher order terms results in

$$\begin{aligned} h(\tilde{x}) &\doteq h(\tilde{x} - \zeta) + \zeta h'(\tilde{x}), \\ h(\hat{x}) &\doteq h(\hat{x} + \delta) - \delta h'(\hat{x}). \end{aligned}$$

Multiplying the above two equations by γ and $1 - \gamma$, respectively, and rearranging terms give

$$\begin{aligned} \gamma h(\tilde{x}) &\doteq \gamma \left(h(\tilde{x} - \zeta) + \frac{\zeta h'(\tilde{x})}{h(\tilde{x} - \zeta)} h(\tilde{x} - \zeta) \right), \\ (1 - \gamma) h(\hat{x}) &\doteq (1 - \gamma) \left(h(\hat{x} + \delta) - \frac{\delta h'(\hat{x})}{h(\hat{x} + \delta)} h(\hat{x} + \delta) \right). \end{aligned}$$

Now, for any sufficiently small $\zeta > 0$, one can pick $\delta > 0$ small enough, such that

$$-\frac{\gamma \zeta h'(\tilde{x})}{h(\tilde{x} - \zeta)} = -\frac{(1 - \gamma) \delta h'(\hat{x})}{h(\hat{x} + \delta)} \triangleq \gamma_0, \quad (10)$$

and $\gamma_0 > 0$, and we then have

$$\begin{aligned} \gamma h(\tilde{x}) &\doteq (\gamma - \gamma_0) h(\tilde{x} - \zeta), \\ (1 - \gamma) h(\hat{x}) &\doteq (1 - \gamma + \gamma_0) h(\hat{x} + \delta), \end{aligned}$$

which means that by deleting γ_0 from γ at position $\tilde{x} - \zeta$ and adding γ_0 to $1 - \gamma$ at position $\hat{x} + \delta$, one can keep the entropy constraint satisfied. Now, denote by $\Lambda_{u'}(z)$ the result after the above operations. Then, we can obtain

$$\begin{aligned}\Lambda_{u'}(z) - \Lambda_u(z) &= (\gamma - \gamma_0)(1 - z) + (1 - \gamma + \gamma_0)(1 - \hat{x} - \delta) \\ &\quad - \gamma(1 - z) - (1 - \gamma)(1 - \hat{x}) \\ &= \gamma_0(z - \hat{x}) - (1 - \gamma + \gamma_0)\delta < 0,\end{aligned}$$

which shows that $\Lambda_u(z)$ is not minimal at z , contradicting the assumption that the channel u achieves the minimum value $\underline{\Lambda}(z)$ at z . ■

Lemma 9 shows that, for the channel u achieving $\underline{\Lambda}(z)$ at z , its probability masses and in particular their associated positions cannot be arbitrary. Indeed, only the following three cases are possible.

- (i) There is only one probability mass on the interval $[0, 1]$. Then the channel u is in fact a BSC such that there is a probability mass 1 at position $1 - 2\epsilon_{\text{bsc}}$, and $\Lambda_{\text{bsc}}(z) = 2\epsilon_{\text{bsc}}$ if $z \in [0, 1 - 2\epsilon_{\text{bsc}})$, while $\Lambda_{\text{bsc}}(z) = 1 - z$ if $z \in [1 - 2\epsilon_{\text{bsc}}, 1]$.
- (ii) There are two probability masses γ and $1 - \gamma$ at positions \tilde{x} and \hat{x} , respectively. Particularly, $\tilde{x} = 0$, and $\hat{x} \in [1 - 2\epsilon_{\text{bsc}}, 1)$ and $\hat{x} \geq z$. In this case, we have $\Lambda_u(z) = \gamma(1 - z) + (1 - \gamma)(1 - \hat{x})$, and the entropy constraint is $\gamma + (1 - \gamma)\mathbf{h}(\hat{x}) = 1 - c$. Notice that \tilde{x} and γ are parameters that should be optimized. Denote by $\Lambda_{\text{opt}}(z)$ the corresponding optimal value.
- (iii) There are two probability masses $1 - c$ and c at positions $\tilde{x} = 0$ and $\hat{x} = 1$; namely, the channel u is a BEC, and $\Lambda_{\text{bec}} = (1 - c)(1 - z)$ for $z \in [0, 1]$.

In order to find the minimum value $\underline{\Lambda}(z)$ at each point $z \in [0, 1]$, now it suffices to compare the point-wise results of the above three cases. Case (ii) above is a non-trivial case, since \hat{x} and γ are unknown parameters. However, the next lemma characterizes the optimal solutions of these parameters.

Lemma 10. The optimization problem

$$\begin{aligned}\underset{\gamma, \hat{x}}{\text{minimize}} \quad & \gamma(1 - z) + (1 - \gamma)(1 - \hat{x}), \\ \text{subject to} \quad & \gamma + (1 - \gamma)\mathbf{h}(\hat{x}) = 1 - c.\end{aligned}\tag{11}$$

has the optimal solutions $\gamma(z) = (1 - c - \mathbf{h}(x(z)))/(1 - \mathbf{h}(x(z)))$ and $x(z)$ satisfying the fixed-point equation $x(z) = (1 - x(z))^{\frac{z-1}{z+1}} - 1$.

Proof: See Appendix A for a complete proof. ■

The next theorem characterizes the exact formula of $\underline{\Lambda}(z)$.

Theorem 11. Consider the channel family $\{\text{BMS}(c)\}$. Then,

$$\underline{\Lambda}(z) = \begin{cases} 2\epsilon_{\text{bsc}} & \text{if } 0 \leq z < z_{\text{bsc}}, \\ \frac{1 - c - \mathbf{h}(x(z))}{1 - \mathbf{h}(x(z))}(1 - z) \\ \quad + \frac{c}{1 - \mathbf{h}(x(z))}(1 - x(z)) & \text{if } z_{\text{bsc}} \leq z < 1, \\ 0 & \text{if } z = 1, \end{cases}$$

where $x(z)$ is the solution of $x = (1 - x)^{\frac{z-1}{z+1}} - 1$ and

$$z_{\text{bsc}} \triangleq \frac{\log_2(4\epsilon_{\text{bsc}}\bar{\epsilon}_{\text{bsc}})}{\log_2(\epsilon_{\text{bsc}}/\bar{\epsilon}_{\text{bsc}})}.\tag{12}$$

Proof: First, the equation $x(z) = (1 - x(z))^{\frac{z-1}{z+1}} - 1$ for $z \in [0, 1)$ is equivalent to

$$z(x) = \frac{\log_2(1 - x) + \log_2(1 + x)}{\log_2(1 - x) - \log_2(1 + x)},\tag{13}$$

where $\lim_{x \rightarrow 1} z(x) = 1$. Plugging $x = 1 - 2\epsilon_{\text{bsc}}$ into (13) yields (12). It is then clear that, for $z \in [0, z_{\text{bsc}})$,

$$\underline{\Lambda}(z) = \min\{\Lambda_{\text{bsc}}(z), \Lambda_{\text{bec}}(z)\} = 2\epsilon_{\text{bsc}}.$$

Note that, for $z = z_{\text{bsc}}$, we have $\Lambda_{\text{bsc}}(z) = \Lambda_{\text{opt}}(z)$. For $z \in [z_{\text{bsc}}, 1)$, the monotonically non-increasing property of the Λ function shows that $\Lambda_{\text{opt}}(z) \leq \Lambda_{\text{bsc}}(z)$. Moreover, when $x(z) \rightarrow 1$ as $z \rightarrow 1$, the continuity of $\mathbf{h}(\cdot)$ shows that, the negative of the slope of $\Lambda_{\text{opt}}(z)$ is equal to

$$\lim_{x \rightarrow 1} \frac{1 - c - \mathbf{h}(x(z))}{1 - \mathbf{h}(x(z))} = 1 - c,$$

which is equal to the negative of the slope of $\Lambda_{\text{bec}}(z)$. Since the slope of $\Lambda_{\text{opt}}(z)$ at z_{bsc} is 0 and $\Lambda_{\text{bec}}(z)$ has a constant slope, it follows from the convex- \cap property of the Λ function that $\Lambda_{\text{opt}}(z) \leq \Lambda_{\text{bec}}(z)$, for $z \in [z_{\text{bsc}}, 1)$. Consequently, when $z \in [z_{\text{bsc}}, 1)$, we have

$$\begin{aligned}\underline{\Lambda}(z) &= \gamma(z)(1 - z) + (1 - \gamma(z))(1 - x(z)) \\ &= \frac{1 - c - \mathbf{h}(x(z))}{1 - \mathbf{h}(x(z))}(1 - z) + \frac{c}{1 - \mathbf{h}(x(z))}(1 - x(z)).\end{aligned}$$

Trivially, when $z = 1$, $\underline{\Lambda}(z) = 0$, completing the proof. ■

Now denote by $\Lambda_*(z)$ the Λ function of the least upgraded channel with respect to the whole channel family $\{\text{BMS}(c)\}$. Then, it is not difficult to see the following.

Corollary 12. $\Lambda_*(z) = \underline{\Lambda}(z)$, for $z \in [0, 1]$.

Denote by C_* the capacity of the least upgraded channel. Expressing z in terms of x , inserting the formula for $\Lambda_*(x(z))$ into (7), and integrating over x gives C_* . However, there is no simple formula in the closed form for C_* . Instead, we compute it numerically in Section IV.

IV. SIMULATION RESULTS

In this section, we provide several simulation results regarding channel degradation and upgradation. From Fig. 1, one can see that the difference between $\bar{\Lambda}(z)$ and $\Lambda^*(z)$ is on the interval $[0, 1 - 2\epsilon_{\text{bsc}})$. On this region, $\bar{\Lambda}(z)$ is convex- \cup while $\Lambda^*(z)$ is linear and thus convex- \cap .

Theorem 3 and Theorem 11 suggest that channels with very few probability masses always achieve the point-wise maximum or point-wise minimum. Thus, we randomly generate 5,000 BMS channels of capacity 0.5 having 2 probability masses, and 5,000 BMS channels of capacity 0.5 having 3 probability masses. Fig. 2 depicts the simulation results.

Other interesting quantities are the gap between c and the capacity of the least degraded channel, call it d^{gap} , and the gap

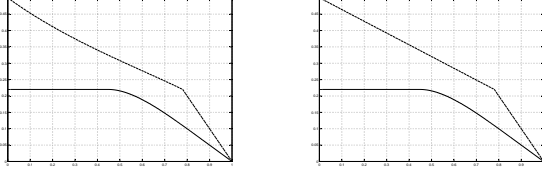


Fig. 1: Left: The point-wise maximum $\bar{\Lambda}(z)$ (dashed line) and the point-wise minimum $\underline{\Lambda}(z)$ (solid line) of $\Lambda_a(z)$, for all $a \in \{\text{BMS}(0.5)\}$. Right: The Λ functions of the least degraded channel $\Lambda^*(z)$ (dashed line) and the least upgraded channel $\Lambda_*(z)$ (solid line), respectively.

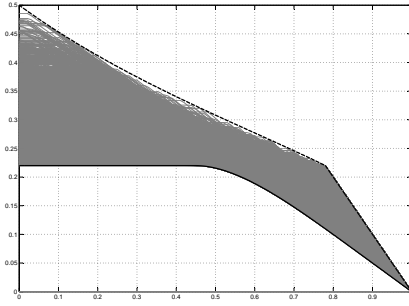


Fig. 2: The Λ functions of randomly generated BMS channels (grey lines) from $\{\text{BMS}(0.5)\}$, and the point-wise maximum $\bar{\Lambda}(z)$ (dashed line) and minimum $\underline{\Lambda}(z)$ (solid line) of $\Lambda_a(z)$, for all $a \in \{\text{BMS}(0.5)\}$.

c	d^{gap}	u^{gap}
0.1	0.0728	0.0686
0.2	0.1222	0.1009
0.3	0.1552	0.1185
0.4	0.1739	0.1257
0.5	0.1795	0.1243
0.6	0.1721	0.1155
0.7	0.1516	0.0995
0.8	0.1175	0.0762
0.9	0.0684	0.0446

TABLE I: The gap between c and C^* is $d^{\text{gap}} = c - C^*$, and the gap between C_* and c is $u^{\text{gap}} = C_* - c$.

between the capacity of the least upgraded channel and c , call it u^{gap} . See Table I for details.

Fig. 3 depicts the capacity C^* (C_* , respectively) of the least degraded channel (resp. the least upgraded channel) as a function of c , where c ranges from 0.001 to 0.999 with a step size of 0.001. We then compute the maximum value of d^{gap} being approximately 0.1795 which corresponds to $c = 0.4940$; while the maximum value of u^{gap} is approximately 0.1261 which corresponds to $c = 0.4310$.

V. CONCLUSION AND OPEN PROBLEMS

The least degraded channel and the least upgraded channel with respect to the family of BMS channels of fixed capacity were introduced in this paper. Also, their characterizations and capacity formulae were derived. An interesting open question is to consider other families of channels. E.g., consider the family of channels which are the result of taken the family of

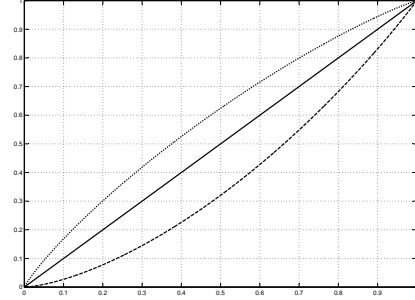


Fig. 3: The channel capacity of the least degraded channel (dashed line) as a function of the underlying channel capacity c , $0 < c < 1$; the channel capacity of the least upgraded channel (dotted line) as a function of the underlying channel capacity c , and the underlying channel capacity (solid line) as a function of itself.

channels of fixed capacity c and performing one polarization step to them. We reserve such questions for future work.

ACKNOWLEDGEMENT

This work was supported by grant No. 200021-125347 of the Swiss National Foundation.

APPENDIX

A. Proof of Lemma 10

Proof: Define $f(\gamma, \hat{x}, \eta) \triangleq \gamma(1 - z) + (1 - \gamma)(1 - \hat{x}) - \eta(\gamma + (1 - \gamma)h(\hat{x}) - 1 + c)$. Then, taking the first-order derivatives with respect to \hat{x} , γ , and η yields

$$\frac{\partial}{\partial \hat{x}} f(\gamma, \hat{x}, \eta) = \gamma - 1 - \eta(1 - \gamma)h'(\hat{x}), \quad (14)$$

$$\frac{\partial}{\partial \gamma} f(\gamma, \hat{x}, \eta) = 1 - z - (1 - \hat{x}) - \eta + \eta h(\hat{x}), \quad (15)$$

$$\frac{\partial}{\partial \eta} f(\gamma, \hat{x}, \eta) = \gamma + (1 - \gamma)h(\hat{x}) - 1 + c. \quad (16)$$

Now setting the first-order derivatives to be 0 gives

$$\gamma = \frac{1 - c - h(\hat{x})}{1 - h(\hat{x})} \quad \text{and} \quad \eta = \frac{\hat{x} - z}{1 - h(\hat{x})}. \quad (17)$$

Plugging (17) into (14) and making some manipulations give

$$(1 + z) \log_2(1 + \hat{x}) + (1 - z) \log_2(1 - \hat{x}) = 0,$$

which is equivalent to

$$\hat{x} = (1 - \hat{x})^{\frac{z-1}{z+1}} - 1. \quad \blacksquare$$

REFERENCES

- [1] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, UK: Cambridge University Press, 2008.
- [2] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3051-3073, 2009.